Data Ordering Genetic Optimization (DOGO) – A Data-Driven Quality Estimate for Every Observation

Dr. Lukas Mandrake, Masha Liukis, Steven Lu, and James Montgomery
Jet Propulsion Laboratory, California Institute of Technology

Poster ID 41, Second AI and Data Science Workshop for Earth and Space Sciences, February 11th, 2021

1. Quality Flags

- Space-based data sets (e.g., OCO-2, OCO-3) often contain quality flags.
- Guide users to find data to use for their analyses.
- Quality flags are great utility, but have drawbacks:
  - One-time optimization – not customized for your analysis
  - Assumes data is good or bad – data quality is not Yes/No; throw away too much data

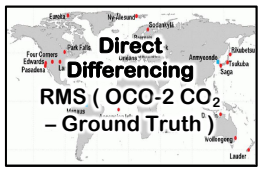
2. Instead of Flags, Order the Data

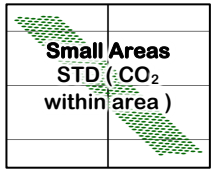
- No good or bad decisions, cutoffs, and lost data
- User specifies how much data from best to worst
- Tunable filter specific for every analysis
- Specify filtration strategy by a single threshold value
- Reproducible results, more comparable findings

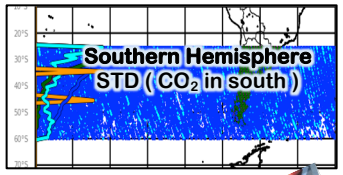
3. Data Driven – Data Ordering Genetic Optimization

Turn the objections into statistical metrics to optimize

- Minimize MEAN(MONTHLY(STDEV(CO2))) in south
- Minimize MEAN(STDEV(CO2)) at small spatial scales
- Minimize RMS(CO2 – ground_truth_CO2)







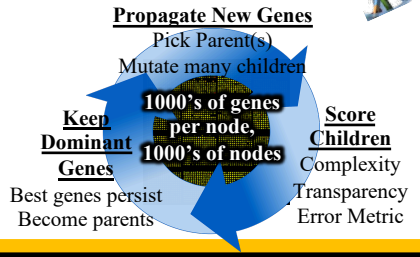
Gene = Data Quality Flag

- Define the gene
- Judge the gene
- Gene metrics

Field	Lower Limit (> or =)	Upper Limit (< or =)
Outcome flag (not in lite file)	N/A	2
Preprocessors/h2o_ratio	0.700	1.030
Preprocessors/co2_ratio	0.995	1.025
Preprocessors/dp_apb	-15.00	5.00
Retrieval/dp	-5.00	10.0
Retrieval/aod_ice	N/A	0.050
Retrieval/Aod_sulfate	N/A	0.400
Retrieval/Aod_dust*	0.001	0.30
Retrieval/Co2_grad_del	-70.0	70.0
Retrieval/albedo_2	0.10	N/A

Genetic Optimization

- Optimize a function
- Large compute resources
- Handle poorly behaved data that may be noisy



4. Results - Warn Levels

Every Observation Gets Its Judgement

- DOGO produces optimal quality flags for every 10% data accepted
- Each observation is examined: how many quality flags would reject it? 0 rejection => Warn Level 0
- Warn Levels are officially delivered to OCO-2 & OCO-3 user community
- Warn Levels are explainable

5. Future Work

- Expand DOGO to support more missions
- Simplify DOGO interface for easy use

6. Acknowledgement

The authors want to thank the OCO-2 & OCO-3 missions and Multi-Mission Ground System and Services (MGSS) for the continuous support of DOGO development.

Contact: lukas.mandrake@jpl.nasa.gov and you.lu@jpl.nasa.gov